Z. Wahrscheinlichkeitstheorie verw. Gebiete
61, 467–481 (1982)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Necessary and Sufficient Conditions for the Pointwise Convergence of Nearest Neighbor Regression Function Estimates

Luc Devroye*

McGill University, School of Computer Science, 805 Sherbrooke Str. West,
Montreal, Canada H3A 2K6

1. Introduction

Let $(X, Y), (X_1, Y_1), \ldots, (X_n, Y_n)$ be independent identically distributed $R^d x[-c, c]$-valued random vectors, and let $m(x)=E(Y|X=x)$ be the *regression function* of Y on X that has to be estimated from the *data* $(X_1, Y_1), \ldots, (X_n, Y_n)$. The *nearest neighbor estimate* is defined by

$$m_n(x)=\sum_{i=1}^{n} v_{ni} Y_i(x), \tag{1}$$

where $(v_{n1}, \ldots, v_{nn})$ is a given probability vector, and $(X_1(x), Y_1(x)), \ldots, (X_n(x), Y_n(x))$ is a permutation of $(X_1, Y_1), \ldots, (X_n, Y_n)$ according to increasing values of $\|X_i - x\|$, $x \in R^d$. When $\|X_i - x\| = \|X_j - x\|$ but $i<j$, X_i is said to be closer to x than X_j. The consistency properties of m_n for special choices of the *weight vector* $(v_{n1}, \ldots, v_{nn})$ are discussed in Cover (1968), Stone (1977), Devroye (1978) and Collomb (1979, 1980). For an analysis of the bias and variance with rate of convergence results, see Lai (1977) and Mack (1981). See also the survey by Collomb (1981). In this paper we give necessary and sufficient conditions on the weight vector for weak, strong and complete pointwise convergence of m_n to m under no assumptions whatsoever on the probability measure μ of X.

Any Borel measurable function of x and the data will be called a regression function estimate. We let $\mathscr{A}$ be the collection of all random vectors (X, Y) taking values in $R^d x[-c, c]$ for some integer $d \geqq 1$ and some constant $c \geqq 0$.

Definition. A regression function estimate m_n is *wpc* (*w*eakly *p*ointwise *c*onsistent) if for all (X, Y) in $\mathscr{A}$,

$$m_n(x) \to m(x) \text{ in probability as } n \to \infty, \text{ almost all } x(\mu). \tag{2}$$

It is *spc* (*s*trongly *p*ointwise *c*onsistent) if in (2) "in probability" can be replaced by "almost surely". It is *cpc* (*c*ompletely *p*ointwise *c*onsistent) if in (2)

* This research was sponsored by National Research Council of Canada Grant No. A 3456

0044-3719/82/0061/0467/$03.00

"in probability" can be replaced by "completely" (for the definition of complete convergence, see Stout (1974, pp. 255)).

From the pointwise convergence of m_n, one can often deduce results about the convergence of $\int |m_n(x)-m(x)|^q \mu(dx)$ ($q \geqq 1$) but the inverse deduction is not simple. Thus, our results cannot be used directly to obtain necessary and sufficient conditions for the integral convergence of m_n.

In Sect. 2, several lemmas of independent interest are stated. The necessary and sufficient conditions for the properties "wpc", "spc" and "cpc" of the nearest neighbor estimate are given in Sect. 3.

2. Lemmas

Lemma 1. *(Binomial tail probabilities.) Let $p \in (0, \frac{1}{2})$ and $n \geqq 1$ be given; p may depend upon n. Let $b(i,n,p) = \binom{n}{i} p^i (1-p)^{n-i}$ be the i-th binomial probability, and let $B(k,n,p) = \sum_{i=0}^{k} b(i,n,p)$. If k, and p vary with n in such a way that $k \to \infty$, $k^2/n \to 0$ and $k/(np) \to 0$, then*

$$B(k,n,p) = o(e^{-(1+o(1))np}).$$

Also, when $(np)/\log n \to \infty$, then

$$\sum_{n=1}^{\infty} B(k,n,p) < \infty.$$

Proof of Lemma 1. Check that

$$B(k,n,p) \sim b(k,n,p) \sim \left(\frac{nep}{k(1-p)}\right)^k (1-p)^n (2\pi k)^{-1/2}.$$

Lemma 2. *If $0 \leqq a \leqq 1$ and $0 < c$ are constants, then any $[0,c]$-valued random variable X satisfies*

$$P(X \geqq aE(X)) \geqq \frac{1-a}{c} E(X).$$

Proof of Lemma 2. When I is the indicator function, then $E(X) = E(X I_{[X < aE(X)]} + E(X I_{[X \geqq aE(X)]} \leqq aE(X) + cP(X \geqq aE(X))$.

Lemma 3. *If $a_1 \geqq \ldots \geqq a_n \geqq 0$ and $b_1, \ldots, b_n$ are real numbers, then*

$$\left|\sum_{i=1}^{n} a_i b_i\right| \leqq a_1 \sup_{i \leqq n} \left|\sum_{j=1}^{i} b_j\right|.$$

Proof of Lemma 3.

$$\left|\sum_{i=1}^{n} a_i b_i\right| = \left|\sum_{i=1}^{n} \left(\sum_{j=1}^{i} b_j\right)(a_i - a_{i+1})\right| \quad \text{(where } a_{n+1}=0)$$
$$\leqq \sum_{i=1}^{n} \left|\sum_{j=1}^{i} b_j\right| (a_i - a_{i+1})$$
$$\leqq \sup_{i \leqq n} \left|\sum_{j=1}^{i} b_j\right| a_1.$$

Lemma 4. *(The complete convergence of $X_k(x)$ to x.) Let $x \in \mathrm{support}(\mu)$, and let $X_k(x)$ be the k-th nearest neighbor of x among $X_1, \ldots, X_n$. Then $k/n \to 0$ as $n \to \infty$ implies that $\|X_k(x) - x\| \to 0$ completely as $n \to \infty$.*

Proof of Lemma 4. Let $\varepsilon > 0$ be arbitrary, and let $p = P(\|X - x\| \leqq \varepsilon)$. Clearly, $p > 0$. If Z is a binomial (n, p) random variable, then for all n large enough,

$$P(\|X_k(x) - x\| > \varepsilon) \leqq P(Z < k) \leqq P\left(Z - np < -\frac{np}{2}\right) \leqq e^{-2n(p/2)^2} = e^{-np^2/2}.$$

Here we used Hoeffding's inequality (Hoeffding, 1963). These probabilities are summable in n for all $\varepsilon > 0$.

Lemma 5. *(An extension of Kolmogorov's exponential inequalities.)*

Let $Y_1, \ldots, Y_n$ be i.i.d. nondegenerate random variables. Let $a_1, \ldots, a_n$ be nonnegative numbers such that

(i) $\sum_{i=1}^{n} a_i \leqq 1$;

(ii) *there exists $b > 0$ such that $a = \sum_{i=1}^{n} a_i^2 \geqq b \sup_i a_i$.*

Then there exist constants c_1, c_2, $\sigma > 0$ independent of $a_1, \ldots, a_n$ (but possibly depending upon the distribution of Y_1) such that for all $\varepsilon \in (c_1 \sqrt{a}, c_2 b)$,

$$P\left(\left|\sum_{i=1}^{n} a_i Y_i\right| > \varepsilon\right) \geqq \tfrac{1}{2} \exp\left(-\frac{4\varepsilon^2}{a\sigma^2}\right).$$

Proof of Lemma 5. Let $Y_1', \ldots, Y_n'$ be distributed as and independent of $Y_1, \ldots, Y_n$. Let $\bar{Y}_i'$ and $\bar{Y}_i$ be equal to Y_i' and Y_i truncated at $\pm\delta$ where $\delta > 0$ is chosen such that the variance σ^2 of $\bar{Y}_1 - \bar{Y}_1'$ is nonzero. Exploiting symmetry, we have

$$P\left(\left|\sum_{i=1}^{n} a_i Y_i\right| > \varepsilon\right) \geqq \tfrac{1}{2} P\left(\left|\sum_{i=1}^{n} a_i (Y_i - Y_i')\right| > 2\varepsilon\right)$$
$$\geqq \tfrac{1}{4} P\left(\left|\sum_{i=1}^{n} a_i (\bar{Y}_i - \bar{Y}_i')\right| > 2\varepsilon\right) = \tfrac{1}{2} P\left(\sum_{i=1}^{n} Z_i > 2\varepsilon\right)$$

where $Z_i = a_i(\bar{Y}_i - \bar{Y}_i')$. Note that $\mathrm{Var}(Z_i) = \sigma^2 a_i^2$, $\sum \mathrm{Var}(Z_i) = s_n^2 = \sigma^2 a$, and $|Z_i| \leqq c s_n$ where $c = 2\delta \sup a_i / s_n$. By Kolmogorov's exponential inequalities (see for example Stout (1974, pp. 262)), there exist constants $b_1, b_2 > 0$ such that

$b_1 < \theta < b_2/c$ implies $P(\sum Z_i > \theta s_n) \geqq \exp(-\theta^2)$. Thus, for $\varepsilon \in (b_1 s_n/2, b_2 s_n^2/(4\delta \sup_i a_i))$, we have

$$P\left(\left|\sum_{i=1}^n a_i Y_i\right| > \varepsilon\right) \geqq \tfrac{1}{2}\exp\left(\frac{4\varepsilon^2}{s_n^2}\right) = \tfrac{1}{2}\exp\left(-\frac{4\varepsilon^2}{a\sigma^2}\right).$$

The inequality is valid for all ε in the interval $(b_1 \sigma \sqrt{a}/2, b_2 b\sigma^2/(4\delta))$.

Lemma 6. *(Exponential inequalities for weighted sums.) Let $Y_1, \ldots, Y_n$ be independent zero mean random variables satisfying $|Y_i| \leqq c$ almost surely. Then*

(i) *For all $a_1, \ldots, a_n \geqq 0$ with sum not exceeding 1, and all $\varepsilon > 0$,*

$$P\left(\left|\sum_{i=1}^n a_i Y_i\right| > \varepsilon\right) \leqq 2\exp\left(-\frac{\varepsilon^2}{2(c^2 + c\varepsilon)\sup a_i}\right),$$

and

(ii) *For fixed $a_1 > 0$,*

$$P\left(\sup_{a_1 \geqq a_2 \geqq \cdots \geqq a_n \geqq 0} \left|\sum_{i=1}^n a_i Y_i\right| > \varepsilon\right) \leqq 2\exp\left(-\frac{\varepsilon^2}{2(c^2 n a_1^2 + c\varepsilon a_1)}\right).$$

Proof of Lemma 6. We will use an inequality due to Bennett (1962) and Hoeffding (1963): when $Y_1, \ldots, Y_n$ are independent random variables with zero mean such that $|Y_i| \leqq c$ almost surely, then, for all $\varepsilon > 0$,

$$\begin{aligned} P\left(\left|\frac{1}{n}\sum_{i=1}^n Y_i\right| > \varepsilon\right) &\leqq 2\exp\left(-\frac{n}{2c}\left[\left(1 + \frac{s^2}{2c\varepsilon}\right)\log\left(1 + \frac{2c\varepsilon}{s^2}\right) - 1\right]\right) \\ &\leqq 2\exp\left(-\frac{n\varepsilon^2}{2(s^2 + c\varepsilon)}\right), \end{aligned} \tag{3}$$

where $s^2 = \frac{1}{n}\sum_{i=1}^n \operatorname{Var}(Y_i)$. In the second step, we used the elementary inequality $\frac{2x}{2+x} < \log(1+x)$, $x > 0$. To obtain, (i), apply (3) with ε replaced by $\frac{\varepsilon}{n}$, c replaced by $c \sup a_i$ and s^2 replaced by $\frac{1}{n}\sum a_i^2 \operatorname{Var}(Y_i) \leqq \frac{c^2}{n}\sup a_i$.

Next, by Lemma 3, for fixed $a_1 > 0$, $\varepsilon > 0$,

$$P\left(\sup_{a_1 \geqq a_2 \geqq \cdots \geqq a_n \geqq 0} \left|\sum_{i=1}^n a_i Y_i\right| > \varepsilon\right) \leqq P\left(\sup_{i \leqq n}\left|\sum_{j=1}^i Y_j\right| > \frac{\varepsilon}{a_1}\right). \tag{4}$$

Bennett's inequality $\left(\text{with } \frac{\varepsilon}{n a_1} \text{ instead of } \varepsilon\right)$ is applicable to the right-hand side of (4) (see for example Steiger (1967); or combine Fuk and Nagaev (1971, expression (43)) with Borokov's theorem (Borokov, 1972)). This yields (ii) without further work.

Lemma 7. *Let $X_1, \ldots, X_n$ be i.i.d. uniform $(0, 1)$ random variables, let $a > 2$ be a constant, let n_j be the largest integer in $\exp(aj\log j)$, and let k_n be a sequence of integers such that $k_n \leqq M \log\log n$, $n \geqq 8$, some $M < \infty$. Then, if X_j^* is the smallest order statistic of $X_1, \ldots, X_{n_j}$, and X_j' is the k_{n_j}-th smallest order statistic of*

$X_1, \ldots, X_{n_j}$, *it follows that*

$$X_j' \geqq X_{j-1}^* \text{ finitely often with probability one.}$$

Proof of Lemma 7. It is known that almost surely $X_{j-1}^* < 1/(n_{j-1} \log^2 n_{j-1})$ finitely often (Geffroy (1958); see also Barndorff-Nielsen (1961) or Kiefer (1970)). Thus it suffices to show that

$$P\left(X_j' \geqq \frac{1}{n_{j-1} \log^2 n_{j-1}} \text{ f.o.}\right) = 1.$$

But

$$\begin{aligned} P\left(X_j' \geqq \frac{1}{n_{j-1} \log^2 n_{j-1}}\right) &\leqq B\left(k_{n_j} - 1, n_j, \frac{1}{n_{j-1} \log^2 n_{j-1}}\right) \\ &\leqq B\left(l_{n_j} - 1, n_j, \frac{1}{n_{j-1} \log^2 n_{j-1}}\right) \end{aligned} \tag{5}$$

where B is the binomial tail defined in Lemma 1, and l_n = largest integer in $M \log\log n$. Lemma 7 now follows from the Borel-Cantelli lemma if we can show that the right-hand side of (5) is summable in j.

We note first the following facts:

$$\frac{n_j}{n_{j-1}} \geqq \frac{\exp(aj \log j) - 1}{\exp(a(j-1)\log(j-1))} \sim \exp\left[aj \log \frac{j}{j-1} + a \log(j-1)\right] \geqq (j-1)^a e^a;$$

$$\log^2 n_{j-1} \leqq (aj \log j)^2;$$

$$l_{n_j} \leqq M \log\log n_j \leqq M \log(aj \log j) + o(1) \leqq 2M \log j \qquad \text{for all } j \text{ large enough.}$$

Since $l_{n_j}^2/n_j \to 0$, $l_{n_j} \to \infty$ as $j \to \infty$ and $\dfrac{n_j}{(n_{j-1} \log^2 n_{j-1})(l_{n_j} - 1)} \to \infty$ for all $a > 2$, Lemma 1 is applicable to the right-hand side of (5); the j-th term is

$$o\left(\exp\left(-(1+o(1)) \frac{n_j}{n_{j-1} \log^2 n_{j-1}}\right)\right). \tag{6}$$

Now, $n_j/(n_{j-1} \log^2 n_{j-1}) \geqq (1+o(1))(j-1)^a e^a/(a^2 j^2 \log^2 j)$. For $a > 2$, the terms (6) are summable in j, which concludes the proof of Lemma 7.

Definition. A sequence of nonnegative numbers a_n is said to be *semimonotone* if there exists a $c > 0$ such that for all $n, m \geqq 1$ $a_{n+m} \geqq c a_n$.

We note here that for any semimonotone sequence, either $\limsup a_n < \infty$ or $\lim a_n = \infty$. Also, if b_n is another sequence such that b_n/a_n stays bounded away from 0 and ∞, and a_n is semimonotone, then b_n is semimonotone.

We now present a Lemma regarding sequences of probability vectors $(v_{n1}, \ldots, v_{nn})$.

Lemma 8. 1. *The following conditions are equivalent:*

(A) $$\sum_{i > \varepsilon n} v_{ni} \to 0 \quad \text{as } n \to \infty, \text{ all } \varepsilon > 0.$$

(B) *There exists a sequence of integers* k_n *such that as* $n\to\infty$,

$$k_n\to\infty,\ k_n/n\to 0 \quad \text{and} \quad \sum_{k_n+1}^{n} v_{ni}\to 0.$$

2. *If there exists a positive constant* α *such that* $\sum_{i>\alpha/\sup_i v_{ni}} v_{ni}\to 0$ *as* $n\to\infty$, *then* $\sum_{k_n+1}^{n} v_{ni}\to 0$ *as* $n\to\infty$, *where* $k_n=\operatorname{int}(\alpha/\sup_i v_{ni})$. *If in addition* (A) *holds, then* $k_n/n\to 0$ *and* $n\sup_i v_{ni}\to\infty$ *as* $n\to\infty$. *If* $\sup_i v_{ni}$ *is monotone in n, so is* k_n. *Finally,* $\sum_{i=1}^{n} v_{ni}^2\geqq a\sup_i v_{ni}$ *for some* $a\in(0,1]$ *and all n large enough.*

Proof of Lemma 8. (B) implies (A) since for each $\varepsilon>0$, and all n large enough, $k_n+1<\varepsilon n$. Also, (A) implies (B) by construction: let n_j, $j\geqq 1$ be a sequence of integers such that $1=n_1<n_2\ldots$ and

$$\sum_{i>n/j} v_{ni}<\frac{1}{j}, \qquad \text{all } n\geqq n_j.$$

Let $k_n=j$ on $[n_j,n_{j+1})$. Clearly, $k_n\to\infty$ and $k_n/n\to 0$ as $n\to\infty$. Also, $\sum_{k_n+1}^{n} v_{ni}\to 0$ as $n\to\infty$.

The first statement of part 2 is trivially true. The second statement is valid because for all $\varepsilon>0$, $\varepsilon n\sup_i v_{ni}\geqq\sum_{i\leqq\varepsilon n} v_{ni}\to 1$ as $n\to\infty$. The last statement of part 2 can be shown by using Schwartz's inequality:

$$\sum_{i=1}^{n} v_{ni}^2\geqq\sum_{i=1}^{k_n} v_{ni}^2\geqq\frac{1}{k_n}\left[\sum_{i=1}^{k_n} v_{ni}\right]^2\geqq\frac{1}{2k_n}\geqq\frac{1}{2\alpha}\sup_i v_{ni},$$

valid for all n large enough.

3. Main Results

One or more of the following conditions will be used in this section:

$$\sum_{i=1}^{n} v_{ni}^2\geqq a\sup_i v_{ni}, \qquad \text{some } a>0, \text{ all } n \text{ large enough}; \tag{7}$$

there exists a positive constant α such that

$$\sum_{i>\alpha/\sup_i v_{ni}} v_{ni}\to 0 \qquad \text{as } n\to\infty; \tag{8}$$

$$\sum_{i>\varepsilon n} v_{ni}\to 0 \qquad \text{as } n\to\infty, \text{ all } \varepsilon>0; \tag{9}$$

$$\sup_i v_{ni}\to 0 \qquad \text{as } n\to\infty. \tag{10}$$

Theorem 1. *The nearest neighbor estimate is wpc. when* (8)–(10) *hold. When the nearest neighbor estimate is wpc., then* (9)–(10) *must be satisfied.*

Theorem 2. *The nearest neighbor estimate is cpc. when* (8)–(9) *and*

$$\sup_i v_{ni} \log n \to 0 \quad \text{as } n \to \infty \tag{11}$$

hold. When it is cpc, then (9)–(10) *must be satisfied. Moreover, if* $1/(\sup_i v_{ni} \log n)$ *is semimonotone and* (7) *holds, then* (11) *must be satisfied too.*

Theorem 3. *When the nearest neighbor estimate is spc, then* (9)–(10) *must be satisfied. Moreover, if* (8) *holds, and* $1/(\sup_i v_{ni} \log\log n)$ *is semimonotone, then*

$$\sup_i v_{ni} \log\log n \to 0 \quad \text{as } n \to \infty. \tag{12}$$

Conversely, the nearest neighbor estimate is spc. when (8)–(10) *hold, the convergence in* (10) *is monotone,* (12) *is satisfied,*

$$v_{n1} \geqq v_{n2} \geqq \dots \geqq v_{nn}, \quad \text{all } n, \tag{13}$$

and

$\sup_i v_{ni}$ *is dominatedly varying (i.e., there exists a finite constant* β *such that for all n,* $\sup_i v_{\lfloor n/2 \rfloor i} \leqq \beta \sup_i v_{ni}$*).* (14)

Remark 1. (The k_n-nearest neighbor estimate.)

When $v_{ni} = 1/k_n$, $1 \leqq i \leqq k_n$, and $v_{ni} = 0$, $i > k_n$, where k_n is an integer not exceeding n, then (7), (8) and (13) are satisfied. The theorems given above can be summarized as follows:

1. The estimate is *wpc* if and only if $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$.

2. The estimate is *cpc* if and only if $k_n/\log n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$. For the necessity, we also require that $k_n/\log n$ be semimonotone.

3. The estimate is *spc* if and only if $k_n/\log\log n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$. For the necessity, we also require that $k_n/\log\log n$ be semimonotone. For the sufficiency, we need the additional conditions that k_n is monotone and that there exists a finite constant β such that $k_n \leqq \beta k_{\lfloor n/2 \rfloor}$, all n.

Proof of Theorems 1 and 2

The sufficiency.

Note that $|m_n(x) - m(x)| \leqq U_n(x) + V_n(x)$ where $U_n(x) = \left| \sum_{i=1}^n v_{ni}(Y_i(x) - m(X_i(x))) \right|$ and $V_n(x) = \left| \sum_{i=1}^n v_{ni}(m(X_i(x)) - m(x)) \right|$.

For all x, given $X_1(x), \dots, X_n(x)$, the random variables $Y_i(x) - m(X_i(x))$ are independent zero mean and bounded random variables. Thus, by Lemma 6, for all $\varepsilon>0$,

$$P(U_n(x)>\varepsilon)\leqq 2\exp(-\delta/\sup_i v_{ni}) \tag{15}$$

where $\delta>0$ does not depend upon n or x. The right-hand-side of (15) tends to 0 when (10) is valid. The terms on the right-hand-side of (15) are summable in n when (11) holds.

It is known that for almost all $x(\mu)$

$$\lim_{r\downarrow 0}\int_{\|y-x\|\leqq r}|m(y)-m(x)|\,\mu(dy)\Big/\int_{\|y-x\|\leqq r}\mu(dy)=0 \tag{16}$$

(see, e.g., Wheeden and Zygmund (1977, pp. 189)). For a given version of m, let us call the set on which (16) holds A. Define further $k_n=\operatorname{int}(\alpha/\sup_i v_{ni})$ where α is the constant in (8). For arbitrary $\delta>0$,

$$V_n(x)\leqq 2c\sum_{i=k_n+1}^{n} v_{ni}+2c\,I_{[\|X_{k_n+1}(x)-x\|>\delta]}+V_n'(x) \tag{17}$$

where c is the uniform bound on $|m|$, I is the indicator function of an event, and

$$\begin{aligned}V_n'(x)&=\sum_{i=1}^{k_n} v_{ni}|m(X_i(x))-m(x)|\,I_{[\|X_{k_n+1}(x)-x\|\leqq\delta]}\\ &\leqq V_n''(x)=\frac{\alpha}{k_n}\sum_{i=1}^{k_n}|m(X_i(x))-m(x)|\,I_{[\|X_{k_n+1}(x)-x\|\leqq\delta]}.\end{aligned} \tag{18}$$

By (8)-(9), Lemma 8 and Lemma 4, the first two terms of (17) tend to 0 completely for all $x\in\operatorname{support}(\mu)$. Consider now $V_n''(x)$ for $x\in A\cap\operatorname{support}(\mu)$. If μ were absolutely continuous with respect to Lebesgue measure, this random variable would be easy to deal with. However, when $\|X_i-x\|=\|X_j-x\|$ with positive probability, we must be a bit more careful. Let us artificially attach independent uniform $(0,1)$ random variables $W_1,\dots,W_n$ to $(X_1,Y_1),\dots,(X_n,Y_n)$, and break ties $\|X_i-x\|=\|X_j-x\|$ by comparing the values of W_i and W_j. Clearly, the distribution of $V_n'(x)$ is not affected by this new method of breaking ties. Also, by the probability integral transform,

$$\int_{\|y-x\|<\|X_1-x\|}\mu(dy)+W_1\int_{\|y-x\|=\|X_1-x\|}\mu(dy)$$

is uniformly distributed on $(0,1)$. Let $(x_0,w_0)\in R^d x[0,1]$, and let S be the open sphere centered at x with radius $\|x-x_0\|$. Let C be shell of this open sphere (closure $(S)-S$). By choice of δ,

$$\begin{aligned}&\sup_{\substack{\|x_0-x\|\leqq\delta\\ 0\leqq w_0\leqq 1}}\Big[\int_S|m(y)-m(x)|\,\mu(dy)+w_0\int_C|m(y)-m(x)|\,\mu(dy)\Big]\Big/[\mu(S)+w_0\mu(C)]\\ &=\sup_{\|x_0-x\|\leqq\delta}\max\left[\frac{\int_S|m(y)-m(x)|\,\mu(dy)}{\mu(S)},\frac{\int_{SUC}|m(y)-m(x)|\,\mu(dy)}{\mu(SUC)}\right]\\ &\leqq\frac{\varepsilon}{2\alpha},\quad\text{where }\varepsilon>0\text{ is a given number.}\end{aligned} \tag{19}$$

Conditional on $(X_{k_n+1}(x), W_{k_n+1}(x))=(x_0, w_0)$, we can consider $(X_1(x), W_1(x)), \ldots, (X_{k_n}(x), W_{k_n}(x))$ as an ordered sample from the distribution of (X_1, W_1) restricted to $\|X_1-x\|<\|x_0-x\|$ or $\|X_1-x\|=\|x_0-x\|$, $W_1<w_0$. Let $(X'_1, W'_1), \ldots$ be i.i.d. random variables from this distribution truncated at (x_0, w_0) in the said manner. For arbitrary $\varepsilon>0$, we have

$$P(V''_n(x)>\varepsilon \,|\, (X_{k_n+1}(x), W_{k_n+1}(x))=(x_0, w_0)) \leqq I_{[\|x_0-x\|<\delta]} P\left(\frac{1}{k_n}\sum_{i=1}^{k_n} |m(X'_i)-m(x)|>\frac{\varepsilon}{\alpha}\right).$$

By (19) and Lemma 6, the last expression is not greater than $2\exp(-\gamma k_n)$ where $\gamma>0$ is a constant depending upon ε, c and α only. Taking expectations yields

$$P(V''_n(x)>\varepsilon)\leqq 2e^{-\gamma k_n}\leqq 2e^{-\gamma\alpha/\sup_i v_{ni}}e^{\gamma}. \tag{20}$$

The terms in (20) tend to 0 as $n\to\infty$ when (10) holds. They are summable in n (for all $\varepsilon>0$) when (11) holds. This concludes the sufficiency part of Theorems 1 and 2.

The Necessity. Consider first the necessity of (10) in Theorem 1: let X_1 have a uniform distribution on $[0,1]$, and let Y_1 be a bounded zero mean random variable independent of X_1. Assume that its variance σ^2 is nonzero. For any $x\in R$, $m_n(x)$ is distributed as $\sum_{i=1}^{n} v_{ni} Y_i$, a random variable with zero mean and variance $\sum_{i=1}^{n} \sigma^2 v_{ni}^2$. This random variable is also uniformly bounded in n. Therefore, $m_n(x)\to 0$ in probability only if $\mathrm{Var}(m_n(x))\to 0$ as $n\to\infty$. But this is equivalent to (10).

We now prove the necessity of (9). Let X_1 be uniform on $[0,1]$, and let $Y_1 = X_1^2$. Take $0<x<\frac{1}{2}$, and define $Z_n=\sum_{i=1}^{n} v_{ni}(X_i^2(x)-x^2)$. Let $W_1, \ldots, W_n$ be i.i.d. random variables, independent of the X_i's, such that $P(W_1=1)=P(W_1=-1)=\frac{1}{2}$. Now,

$$\begin{aligned}
E(Z_n)&=E\left(\sum_{i=1}^{n} v_{ni}(X_i^2(x)-x^2)\right)\\
&\geqq E\left(\sum_{i=1}^{n} v_{ni}(X_i^2(x)-x^2)\, I_{[|X_i(x)-x|<x]}\right)\\
&=E\left(\sum_{i=1}^{n} v_{ni}\left((X_i(0)\,W_i+x)^2-x^2\right) I_{[X_i(0)<2x]}\right)\\
&=E\left(\sum_{i=1}^{n} v_{ni}(X_i^2(0)\,W_i^2+2X_i(0)\,W_i)\, I_{[X_i(0)<2x]}\right)\\
&=E\left(\sum_{i=1}^{n} v_{ni}X_i^2(0)\, I_{[X_i(0)<2x]}\right)\\
&\geqq \sum_{i=1}^{n} v_{ni}\left(2x\frac{i}{n+1}\right)^2
\end{aligned}$$

where we used Jensen's inequality. It suffices now to show that $Z_{n'}$ cannot converge to 0 in probability along a subsequence n' of n when for some $\varepsilon>0$, $\delta>0$,

$$\sum_{i>\varepsilon n'} v_{n'i} \geqq \delta > 0$$

along this subsequence. Indeed,

$$\sum_{i=1}^{n'} i^2 v_{n'i} \geqq \sum_{i \geqq \varepsilon n'+1} i^2 v_{n'i} \geqq (\varepsilon n'+1)^2.$$

So, for $\varepsilon \leqq 1$,

$$E(Z_{n'}) \geqq \frac{4x^2}{(n'+1)^2} (\varepsilon n'+1)^2 \delta \geqq 4x^2 \varepsilon^2 \delta.$$

Also,

$$P(Z_{n'} > 2x^2 \varepsilon^2 \delta) \geqq P(Z_{n'} > E(Z_{n'})/2) \geqq E(Z_{n'})/2 \geqq 2x^2 \varepsilon^2 \delta > 0.$$

Therefore, for all $\varepsilon>0$, $\sum_{i>\varepsilon n}^{n} v_{ni} \to 0$ as $n \to \infty$.

To establish the necessity of (11), let X_1 and Y_1 be as in the first example. Let $v_n = \sum_{i=1}^{n} v_{ni}^2$. By Lemma 5, we know that there exist constants c_1, $c_2>0$ such that for all $\varepsilon \in (c_1 \sqrt{v_n}, c_2)$, all $x \in [0,1]$,

$$P(|m_n(x)| > \varepsilon) \geqq \tfrac{1}{2} \exp\left(-\frac{4\varepsilon^2}{v_n \sigma^2}\right). \tag{21}$$

Here the assumption (7) is required. This lower bound is at least equal to $\frac{1}{2}\exp(-c_3/\sup_i v_{ni})$ where $c_3>0$ is a constant. We know also that $\sup_i v_{ni} \log n$ tends to 0 with n, or stays bounded away from 0 as $n \to \infty$. In the latter case, assuming that $\delta \leqq \sup_i v_{ni} \log n$, all n, we see that $\exp(-c_3/\sup_i v_{ni}) \geqq n^{-c_3/\delta}$. The terms in the lower bound are not summable in n when $c_3 \leqq \delta$, and this leads to a contradiction: indeed, since c_3 is proportional to ε^2, we can make it as small as desired.

Proof of Theorem 3

The Necessity. In view of Theorem 1, it suffices to show the necessity of (12). Let X_1 be uniformly distributed on $[0,1]$ and let Y_1 be independent of X_1, bounded ($|Y_1| \leqq 1$) and nondegenerate ($P(Y_1=0)<1$). Also, $E(Y_1)=0$. Let $x=0$ without loss of generality. Define $Z_n = m_n(0) - m(0) = m_n(0)$. Since $a \geqq \sup_i v_{ni}/v_n \geqq 1$, the semimonotonicity of $(\sup_i v_{ni} \log\log n)^{-1}$ and that of $(v_n \log\log n)^{-1}$ are equivalent. Thus, either (12) holds, or there exists a $\delta>0$ such that $v_n \geqq \delta/\log\log n$ for all n large enough.

We will show that under the latter assumption. $|Z_n| > \varepsilon$ i.o. with probability one for all ε small enough. By condition (8) and the boundedness of $|Y_1|$, it suffices to prove that for all ε small enough, $|Z'_n|>\varepsilon$ i.o. with probability one, where

$$Z'_n=\sum_{j=1}^{k_n} v_{nj} Y_j(0),$$

and $k_n=\operatorname{int}(\alpha/\sup_i v_{ni})$. Let us now inherit the notation of the proof of Theorems 1 and 2, and let n be so large that $\sum_{i=1}^{k_n} v_{ni}\geqq\frac{1}{2}$. Then, as in (21), for some $c_1, c_2, b, \sigma>0$, and for all $\varepsilon\in(c_1\sqrt{v'_n}, c_2 b)$,

$$P(|Z'_n|>\varepsilon)\geqq\tfrac{1}{2}\exp\left(-\frac{4\varepsilon^2}{v'_n\sigma^2}\right)$$

where $v'_n=\sum_{j=1}^{k_n} v_{nj}^2$, $b\leqq\inf_n(v'_n/\sup_i v_{ni})$. Now, $v_n\geqq v'_n=v_n-\sum_{j>k_n} v_{nj}^2\geqq v_n(1-o(1))$ $\geqq v_n/2$, all n large enough; and $v'_n/\sup_i v_{ni}\geqq c_3>0$ for all n large enough. Thus, taking $b=c_3$, we have for all $\varepsilon\in(c_1\sqrt{v_n}, c_2 c_3)$ and all n large,

$$\begin{aligned}P(|Z'_n|>\varepsilon)&\geqq\tfrac{1}{2}\exp\left(-\frac{8\varepsilon^2}{v_n\sigma^2}\right)\geqq\tfrac{1}{2}\exp\left(-\frac{8\varepsilon^2}{\delta\sigma^2}\log\log n\right)\\&=\tfrac{1}{2}(\log n)^{-8\varepsilon^2/(\delta\sigma^2)}.\end{aligned}$$

In particular, if n_j is the largest integer in $\exp(aj\log j)$, where $a>2$ is a constant, then

$$\sum_{j=1}^{\infty} P(|Z'_{n_j}|>\varepsilon)=\infty, \qquad \text{all } \varepsilon>0 \text{ small enough.}$$

We will now show that for such $\varepsilon>0$ and for all N, $P(\bigcup_{j\geqq N}[|Z'_{n_j}|>\varepsilon])=1$, which implies that $|Z'_n|>\varepsilon$ i.o. a.s. Let B_N be the event $\bigcap_{j\geqq N}[X'_j\leqq X^*_{j-1}]$ where $X^*_j(X'_j)$ is the distance of x to its nearest neighbor (k_{n_j}-th nearest neighbor) among $X_1,\dots,X_{n_j}$. Since (8) implies that $k_n\leqq M\log\log n$ for some $M<\infty$, the conditions of Lemma 7 are satisfied. Thus, $X'_j\geqq X^*_{j-1}$ f.o. a.s. In other words, $P(B_N)\to 1$ as $N\to\infty$. On B_N, the random variables Z'_{n_j}, $j\geqq N$, are independent. Because B_N is independent of each individual Z'_{n_j}, we have

$$\begin{aligned}P(\bigcup_{j\geqq N}[|Z'_{n_j}|>\varepsilon])&\geqq P(\bigcup_{j\geqq N}[|Z'_{n_j}|>\varepsilon], B_N)\\&=P(B_N)-P(\bigcap_{j\geqq N}[|Z'_{n_j}|\leqq\varepsilon], B_N)\\&=P(B_N)-P(B_N)\prod_{j=N}^{\infty}P(|Z'_{n_j}|>\varepsilon)\\&\geqq P(B_N)-P(B_N)\exp\left(-\sum_{j=N}^{\infty}P(|Z'_{n_j}|>\varepsilon)\right)\\&=P(B_N)\to 1 \text{ as } N\to\infty.\end{aligned} \tag{22}$$

Since the left-hand-side of (22) is nonincreasing in N, each of its terms must be 1. We have thus obtained a contradiction. Hence, (12) must hold.

The Sufficiency. Let $B \subseteq R^d$ be the set of $x \in \text{support}(\mu)$ for which (16) holds. We recall that $\mu(B)=1$.

Let $k_n = \text{int}(\alpha/\sup_i v_{ni})$, let $\varepsilon > 0$ be arbitrary, and define $v = n/2$. We remark, as in the proof of Theorem 1, that $|m_n(x) - m(x)| \leqq U_n(x) + V_n(x)$. Furthermore, let $Z_{ni} = Y_i(x) - m(X_i(x))$ where $x \in B$ is fixed. Also, let $Z_i = Y_i - m(X_i)$, and note that $E(Z_i | X_i) = 0$ a.s., and that $|Z_i| \leqq c < \infty$ a.s.

Consider first $U_n(x)$. Since $\left| \sum_{i=k_n+1}^{n} v_{ni} Z_{ni} \right| \leqq \frac{\varepsilon}{2}$ for all n large enough, it suffices to show that

$$T_n = \sum_{i=1}^{k_n} v_{ni} Z_{ni} \to 0 \text{ a.s.} \tag{23}$$

We let

$$W_n = \sup_{v \leqq j \leqq n} \left| \sum_{i=1}^{k_j} v_{ji} Z_{ni} \right|,$$

and define the events

$$\begin{aligned} E_n = [&\|X_n - x\| \text{ is among the } k_n \text{ smallest order statistics of} \\ &\|X_1 - x\|, \ldots, \|X_n - x\| \text{ (where we use our tie-breaking} \\ &\text{rule that depends upon the indices)}], \\ A_n = {}&[|W_n| > \varepsilon] \cap E_n, \\ B_n = {}&[|W_n| > \varepsilon] \cap [k_n \neq k_{n-1}], \\ C_N = {}&\bigcup_{i=0}^{\infty} [|W_{N2^i}| > \varepsilon]. \end{aligned}$$

The basic observation is that for N large enough,

$$\bigcup_{n \geqq N} [|T_n| > \varepsilon] \subseteq \bigcup_{n \geqq N} (A_n \cup B_n) \cup C_N. \tag{24}$$

We will show that $P(A_n) + P(B_n)$ is summable in n and $\lim_{N \to \infty} P(C_N) = 0$.

By Lemma 6, the monotonicity of v_{n1}, condition (13), $k_n \leqq \alpha/v_{n1}$, $v_{v1} \leqq c'' v_{n1}$ (condition (14); $c'' > 0$ is a constant), we have

$$P(|W_n| > \varepsilon) \leqq 2 \exp\left(-\frac{\varepsilon^2}{2(c^2 k_n v_{v1}^2 + c \varepsilon v_{v1})}\right) \leqq 2 \exp(-c^*/v_{n1}) \tag{25}$$

where

$$c^* = \varepsilon^2 / (2(c^2 \alpha c'' + c \varepsilon c'')) > 0.$$

By (25) and the definition of A_n,

$$P(A_n) \leqq \frac{k_n}{n} 2 \exp\left(-\frac{c^*}{v_{n1}}\right) \leqq \frac{2\alpha}{n v_{n1}} \exp\left(-\frac{c^*}{v_{n1}}\right).$$

The last expression is a unimodal function of v_{n1} with peak at $v_{n1} = \tilde{c}$. For n so large that $v_{n1} < \tilde{c}$, $v_{n1} \leqq \delta/\log\log n$ (where $\delta > 0$ is to be chosen), we have

$$P(A_n) \leqq \frac{2\alpha \log\log n}{\delta n} \exp\left(-\frac{c^*}{\delta}\log\log n\right) = \frac{2}{\delta}\frac{\log\log n}{n(\log n)^{c^*/\delta}},$$

which is summable in n when we choose $\delta < c^*$.

Again by Lemma 6 and the argument given above,

$$\sum_{n \geqq N} P(B_n) \leqq \sum_{n \geqq N: k_n \neq k_{n-1}} 2\exp\left(-\frac{c^*}{v_{n1}}\right) \leqq \sum_{n \geqq N: k_n \neq k_{n-1}} 2\exp\left(-\frac{c^*}{\alpha}k_n\right)$$
$$\leqq \sum_{j=k_N}^{\infty} 2\exp(-c^* j/\alpha) \to 0 \quad \text{as } N \to \infty.$$

Finally, by the monotonicity of v_{n1}, for N so large that $v_{n1} < \delta/\log\log n$, all $n > \lfloor N/2 \rfloor$, where $0 < \delta < c^*$, we have

$$\sum_{i=0}^{\infty} P(|W_{N2^i}| > \varepsilon) \leqq \sum_{i=0}^{\infty} 2\exp(-c^*/v_{N2^i,1})$$
$$\leqq \sum_{i=0}^{\infty} 2(N2^{i-1})^{-1} \sum_{n=\lfloor N2^{i-1}\rfloor+1}^{N2^i} \exp(-c^*/v_{n1})$$
$$\leqq \sum_{n=\lfloor N/2 \rfloor+1}^{\infty} \frac{4}{n}\exp(-c^*/v_{n1})$$
$$\leqq \sum_{n=\lfloor N/2 \rfloor+1}^{\infty} \frac{4}{n(\log n)^{c^*/\delta}}$$
$$\to 0 \quad \text{as } N \to \infty.$$

Hence $P(C_N) \to 0$ as $N \to \infty$.

Let us now consider $V_n(x)$, and apply the inequalities (17)–(18). In the argument given above for $U_n(x)$, we can replace Z_i by $m(X_i) - m(x)$ and Z_{ni} by $m(X_i(x)) - m(x)$. If $c_1, \ldots, c_5$ are positive constants, then we can also replace W_n by $cI_{F_n^c}$ plus

$$\sup_{v \leqq j \leqq n} \frac{c_1}{k_j} \sum_{i=1}^{k_j} |m(X_i(x)) - m(x)| I_{F_n} \leqq \frac{c_2}{k_n} \sum_{i=1}^{k_n} |m(X_i(x)) - m(x)| I_{F_n} = c_3 V_n''(x) I_{F_n}$$

where

$$F_n = [\|X_{k_v}(x) - x\| \leqq \delta]$$

and $\delta > 0$ is to be chosen. By a slight modification of (20), we see that

$$P(c_3 V_n''(x) I_{F_n} > \varepsilon) \leqq c_4 \exp(-c_5/v_{n1}).$$

The events A_n, B_n, C_N and E_n are defined as above with the replaced W_n, and the remainder of the argument given above for $U_n(x)$ can be inherited. This concludes the proof of Theorem 3.

Remark 2. (Related work.) Our theorems are valid with no conditions on (X, Y) other than the a.s. boundedness of Y. Assuming only the finiteness of $E(|Y|^q)$ $(q \geqq 1)$, Stone (1977) has shown that (9), (10) and (13) are sufficient for the weak convergence to 0 of $\int |m_n(x)-m(x)|^q \mu(dx)$. This is not a pointwise result. Devroye (1981) has shown that $m_n(x) \to m(x)$ in probability as $n \to \infty$, almost all $x(\mu)$, when $E(|Y|) < \infty$, $k_n/n \to 0$, $k_n \to \infty$ and $\sup_i v_{ni} \leq M/k_n$ for some $M < \infty$, where

$$k_n = \max\{j: v_{nj} > 0, v_{ni} = 0, \text{all } i > j\}.$$

Also, when Y is a.s. bounded, and (8)–(9), (11) hold, then $m_n(x) \to m(x)$ completely as $n \to \infty$, almost all $x(\mu)$. This corresponds to the sufficiency part of Theorem 2. For further discussions and generalizations of the latter result, see Gyorfi (1981).

Remark 3. (Global convergence.) Beck (1979) has shown that when X has a density, m has a continuous version and Y is a.s. bounded, then the k_n-nearest neighbor estimate satisfies $\int |m_n(x)-m(x)| \mu(dx) \to 0$ completely as $n \to \infty$, when only $k_n/n \to 0$ and $k_n \to \infty$. This result is profound. It cannot be obtained from our theorems for pointwise convergence. It is also not known at this moment whether Beck's conditions on k_n are sufficient for the global convergence result given above when one just assumes that Y is a.s. bounded.

Remark 4. (Discrimination.) Let the data sequences $(X_1, Y_1), \ldots, (X_n, Y_n)$ be i.i.d. $R^d \times \{1, \ldots, M\}$-valued random vectors, distributed as and independent of (X, Y). In discrimination, one estimates Y by $\hat{Y}$, a Borel measurable function of X and the data sequence. Define $\eta_i(x) = P(Y = i | X = x)$, $1 \leq i \leq M$, $x \in R^d$, and the local Bayes risk $r^*(x) = 1 - \max_i \eta_i(x)$. Assume that all the regression functions η_i are estimated from the data sequence, and that these estimates are $\hat{\eta}_i$. The obvious discrimination method would take $\hat{Y} = s$, where s is one of the indices for which $\max_i \hat{\eta}_i(X)$ is achieved. (How a tie is broken is irrelevant in the present context.) The local probability of error is $r_n(x) = 1 - \eta_{\hat{Y}}(x)$. It is clear that

$$\begin{aligned} 0 \leq r_n(x) - r^*(x) &= \max_i \eta_i(x) - \max_i \hat{\eta}_i(x) + \hat{\eta}_{\hat{Y}}(x) - \eta_{\hat{Y}}(x) \\ &\leq 2 \max_i |\eta_i(x) - \hat{\eta}_i(x)|. \end{aligned} \tag{26}$$

Assume now that we use nearest neighbor estimates $\hat{\eta}_i$ that are obtained by using (1) on the data $(X_1, I_{[Y_1 = i]}), \ldots, (X_n, I_{[Y_n = i]}), \ldots$. Then, by (26), all that is said in Theorems 1–3 about the convergence of m_n to m carries over to the convergence of r_n to r^*. Furthermore, since the probability of error is

$$L_n = P(\hat{Y} \neq Y | X_1, Y_1, \ldots, X_n, Y_n) = \int r_n(x) \mu(dx)$$

and the Bayes probability of error is

$$L^* = \inf_{g: R^d \to \{1, \ldots, M\}} P(g(X) \neq Y) = \int r^*(x) \mu(dx),$$

we have by a generalization of the Lebesgue dominated convergence theorem (Devroye and Wagner, 1980), that $L_n \to L^*$ in probability under the conditions of Theorem 1, and $L_n \to L^*$ a.s. under the conditions of Theorem 3. The latter result improves another result of the author (Devroye, 1981).

Acknowledgement. The author wishes to thank Gérard Collomb and Laszlo Györfi for many useful discussions, and an anonymous referee for pointing out improvements in the paper.

References

1. Barndorff-Nielsen, O.: On the rate of growth of the partial maxima of a sequence of independent identically distributed random variables. Math. Scand. **9**, 383–394 (1961)
2. Beck, J.: The exponential rate of convergence of error for $k_n NN$ nonparametric regression and decision. Problems of Control and Information Theory **8**, 303–311 (1979)
3. Bennett, G.: Probability inequalities for the sum of independent random variables. J. Amer. Statist. Assoc. **57**, 33–45 (1962)
4. Borovkov, A.A.: Notes on inequalities for sums of independent variables. Theory of Probability and its Applications **17**, 556–557 (1972)
5. Chow, Y.S., Teicher, H.: Probability Theory: Independence. Interchangeability, Martingales. New York: Springer 1978
6. Collomb, G.: Estimation de la regression par la méthode des k points les plus proches: propriétés de convergence ponctuelle. Comptes Rendus de l'Academie des Sciences de Paris **289**, 245–247 (1979)
7. Collomb, G.: Estimation de la regression par la méthode des k points les plus proches avec noyau. Lecture Notes in Mathematics **#821**, 159–175. Berlin-Heidelberg-New York: Springer 1980
8. Collomb, G.: Estimation non paramétrique de la regression: revue bibliographique. International Statistical Review **49**, 75–93 (1981)
9. Cover, T.M.: Estimation by the nearest neighbor rule. IEEE Transactions of Information Theory **14**, 50–55 (1968)
10. Devroye, L.: The uniform convergence of nearest neighbor regression function estimators and their application in optimization. IEEE Transactions on Information Theory **24**, 142–151 (1978)
11. Devroye, L.: On the almost everywhere convergence of nonparametric regression function estimators. Ann. Statist. **9**, 1310–1319 (1981)
12. Devroye, L., Wagner, T.J.: On the $L1$ convergence of kernel estimators of regression functions with applications in discrimination. Z. Wahrscheinlichkeitstheorie verw. Gebiete **51**, 15–25 (1980)
13. Fuk, D.K., Nagaev, S.V.: Probability inequalities for sums of independent random variables. Theory of Probability and its Applications **16**, 643–660 (1971)
14. Geffroy, J.: Contribution à la théorie des valeurs extrêmes. Publications de l'Institut de Statistique des Universités de Paris **7**, 37–121 (1958)
15. Györfi, L.: Recent results on nonparametric regression estimate and multiple classification. Problems of Control and Information Theory **10**, 43–52 (1981)
16. Hoeffding, W.: Probability inequalities for sums of bounded random variables. J. Amer. Statist. Assoc. **58**, 13–30 (1963)
17. Kiefer, J.: Iterated logarithm analogues for sample quantiles when $p_n \downarrow 0$. Proceedings Sixth Berkeley Symposium on Mathematical Statistics and Probability **1**, 227–244, University of California Press, 1970
18. Lai, S.L.: Large sample properties of k-nearest neighbor procedures. Ph.D. Dissertation, UCLA, 1977
19. Mack, Y.P.: Local properties of $k-NN$ regression estimates. Manuscript, Department of Statistics, University of Rochester, Rochester, New York, 1981
20. Steiger, W.L.: Some Kolmogoroff-type inequalities for bounded random variables. Biometrika **54**, 641–647 (1967)
21. Stone, C.J.: Consistent nonparametric regression. Ann. Statist. **5**, 595–645 (1977)
22. Stout, W.F.: Almost Sure Convergence. New York: Academic Press 1974
23. Wheeden, R.L., Zygmund, A.: Measure and Integral. New York: Marcel Dekker 1977

Received March 10, 1981; in revised form June 10, 1982